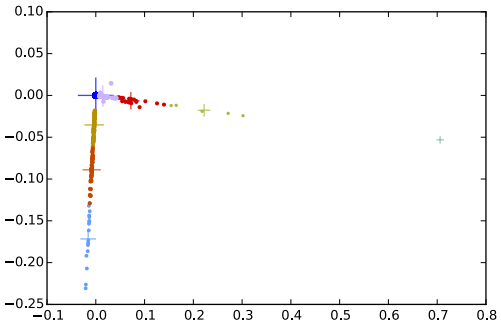
Intelligent Performance Analysis

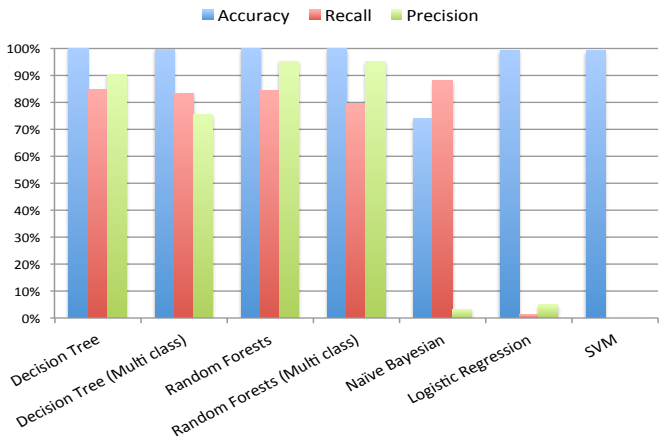
Objectives

- Develop performance analysis techniques for online streaming data.
- Develop online anomaly detection and pattern searching techniques.
- Develop deep learning techniques for performance prediction.



Clustering results of PCA-transformed job log data with L1 norm uniform-scaling to identify performance outliers. Dimensionality reduction can supplement the clustering method when comparing the clustering results and scaling methods. Also, they are useful to identify uncovered performance outliers that clustering method alone cannot detect.

Failure prediction results based on machine learning, 99.8% accuracy with 83.6% recall and 94.8% precision.



Progress & Accomplishments

- Developed automated performance diagnostic method and failure detection method for cluster jobs
- Developed prediction methods for network data throughput and cluster job performance
  - Machine leaning based model (submitted to SC'15)
  - Statistical model - Best Predictive Generalized Linear Mixed Model with Predictive Lasso (IJSP 5/2015 journal paper)
  - Relational Dynamic Bayesian Network (submitted to UAI'15)
  - Time series model (ICNC'15 conf. paper)
- Developed a data reduction method based on pattern searching - U.S. Patent pending serial no. 14/555,365.
- Developing in-situ feature detection methods
- Developing online job failure prediction methods

Impacts

- Improve analysis workflow performance and resource utilization.
  - Enable predictable data throughput over the network.
  - Improve resilience of computing jobs
- Enable dynamic data reduction based on in-situ feature detection

sdmsupport@hpcrd.lbl.gov